CSIRO PUBLISHING
Soil Research, 2014, **52**, 107–119
http://dx.doi.org/10.1071/SR13100

Digital mapping of a soil drainage index for irrigated enterprise suitability in Tasmania, Australia

D. B. Kidd[A,B,C], *B. P. Malone*[B], *A. B. McBratney*[B], *B. Minasny*[B], *and M. A. Webb*[A]

[A]Department of Primary Industries Parks Water and Environment Tasmania, 171 Westbury Road, Prospect, Tas. 7250, Australia.
[B]Faculty of Agriculture and Environment, University of Sydney, 1 Central Avenue, Australian Technology Park, Eveleigh, NSW 2015, Australia.
[C]Corresponding author. Email: darren.kidd@dpipwe.tas.gov.au

Abstract. An operational Digital Soil Assessment was developed to inform land suitability modelling in newly commissioned irrigation schemes in Tasmania, Australia. The Land Suitability model uses various soil parameters, along with other climate and terrain surfaces, to identify suitable areas for various agricultural enterprises for a combined 70 000-ha pilot project area in the Meander and Midlands Regions of Tasmania. An integral consideration for irrigable suitability is soil drainage. Quantitative measurement and mapping can be resource-intensive in time and associated costs, whereas more 'traditional' mapping approaches can be generalised, lacking the detail required for statistically validated products. The project was not sufficiently resourced to undertake replicated field-drainage measurements and relied on expert field drainage estimates at ~930 sites (260 of these for independent validation) to spatially predict soil drainage for both areas using various terrain-based and remotely sensed covariates, using three approaches: (*a*) decision tree spatial modelling of discrete drainage classes; (*b*) regression-tree spatial modelling of a continuous drainage index; (*c*) regression kriging (random-forests with residual-kriging) spatial modelling of a continuous drainage index. Method *b* was chosen as the best approach in terms of interpretation, and model training and validation, with a concordance coefficient of 0.86 and 0.57, respectively. A classified soil drainage map produced from the 'index' showed good agreement, with a linearly weighted kappa coefficient of 0.72 for training, and 0.37 for validation. The index mapping was incorporated into the overall land suitability model and proved an important consideration for the suitability of most enterprises.

Additional keywords: decision trees, digital soil mapping, land suitability, regression trees, random forests, soil drainage, spatial modelling.

Received 25 March 2013, accepted 24 September 2013, published online 6 March 2014

Introduction

Tasmania has a population of ~500 000 people, with a cool temperate climate, and rainfall ranging from >500 mm year^{-1} in the central Midlands to <1800 mm year^{-1} on the West Coast. It has some of the most productive soils in Australia, ranging from the fertile and well-draining Red Ferrosols (Isbell 2002) (Nitisols or Acrisols; IUSS Working Group WRB 2007) in the north-west, to the more poorly drained duplex (prominent change in texture between the A and B horizons) Sodosols (Isbell 2002) (Solonetz or Lixisols; IUSS Working Group WRB 2007) in the drier parts of the state (Cotching and Kidd 2010). Government-commissioned irrigation schemes have been introduced primarily to intensify and diversify agricultural and horticultural production, capitalising on the state's favourable climate and soils to ensure food security and economic prosperity.

The Tasmanian Department of Primary Industries, Parks, Water and Environment (DPIPWE), along with the University of Sydney Faculty of Agriculture and Environment, the Tasmanian Institute of Agriculture (TIA), and the Department of Economic Development, Tourism and the Arts (DEDTA), have developed the 'Wealth from Water' project, which aims to classify land within these schemes for suitability of 20 agricultural enterprises (http://www.dpiw.tas.gov.au/wealthfromwater). The land suitability assessment provides comprehensive soil, climate, and enterprise data, complemented by market and business information (Kidd *et al.* 2012). Completed in late 2012, two irrigation areas were chosen to pilot the suitability and Digital Soil Mapping (DSM) process, namely the Meander Valley of Tasmania and the Tunbridge district of the Midlands Water Scheme, totalling 70 000 ha. Both areas are representative of a range of typical Tasmanian cropping soils and climatic conditions (Fig. 1).

An integral component of any suitability assessment is appropriately detailed soils information. Existing soil mapping for the project areas was not of the scale, format or

Journal compilation © CSIRO 2014

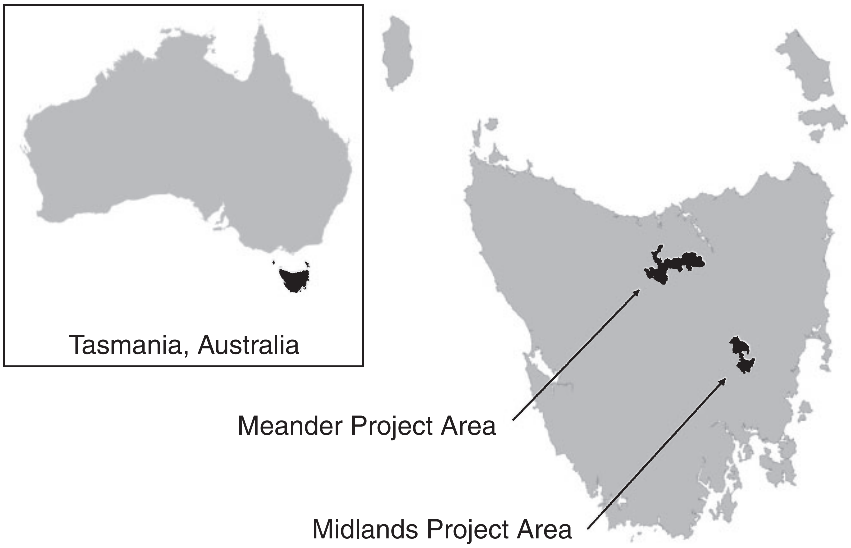

Fig. 1. Project area locations.

quality for the assessment requirements; it consists of 1 : 100 000 vector soil-type mapping undertaken by CSIRO in the 1950s, recently updated by DPIPWE (Leamy 1961; Spanswick and Kidd 2001; Spanswick and Zund 1999). Various DSM techniques were used to generate soil-property maps for the suitability process. A multitude of research and references demonstrates the benefits of DSM and, furthermore, its acceptance into mainstream land resource assessment (Grunwald 2010; McBratney *et al*. 2003). These techniques are now being adopted by various government agencies around Australia into core resource-assessment operations.

Many of the selected enterprises cannot tolerate poorly or rapidly drained soils and require sufficient moisture-holding capacity. Consequently, one of the first and potentially most challenging soil input surfaces developed was soil drainage. Physical drainage capacities within Australia can be quantitatively assessed by field hydraulic property measurements including field saturated hydraulic conductivity (well permeameter, ponded disc infiltrometer, double ring infiltrometer, rainfall simulator), and laboratory measurement of hydraulic conductivity (constant or falling head infiltrometer) (McKenzie *et al*. 2002). These methods can be slow and arduous, requiring significant replication, and are therefore expensive. Due to insufficient time and resources, the project did not undertake replicated field measurements to inform drainage, and required a less resource-intensive alternative. A lack of detailed groundwater measurements in both areas meant that hydrological soil-moisture modelling was not feasible either.

Documented approaches of digital soil drainage mapping include modelling with multi-spectral satellite remote sensing, generation of predictive covariates and simulation techniques. Lemercier *et al*. (2012) used extrapolation of expert soil knowledge (as existing conventional soil mapping) with boosted classification and regression trees, first by predicting soil parent material, then using this as a predictor of natural soil drainage to develop a soil-drainage model. The model was used to extrapolate drainage predictions into surrounding regions and was tested using a validation set with good results. Niang *et al*. (2012) predicted soil drainage based on land-use types from soil survey data and RADAR-satellite imagery as predictors, using both a discriminant analysis and decision tree (DT) classifiers. This approach showed good validation agreement with an existing conventional soil map, and the usefulness of RADAR-satellite remote sensing as a predictor of drainage. Peng *et al*. (2003) also demonstrated the effectiveness of remote sensing analyses of various images to delineate soil drainage classes (limiting this study to bare-earth examples). Bell *et al*. (1992, 1994) used multivariate discriminant analysis to develop a statistical soil–landscape model, then validated drainage-point estimates using class-frequency information. Liu *et al*. (2008) used a combination of electromagnetic induction (EMI) as apparent electrical conductivity, hyperspectral satellite imagery, synthetic aperture RADAR and a GPS-derived digital elevation model (DEM) as predictors for surveyor soil drainage-class field estimates. The study found that a combination of these technologies produced good predictions of spatial drainage classes, improving overall predictions from using DEM terrain products alone. Malone *et al*. (2012) used soil colour as an indicator of external soil drainage in the Hunter Valley region of New South Wales, Australia, where point data of soil colour were used, in conjunction with known colour-based drainage landscape sequences and terrain-based covariates, to generate a continuous soil drainage index, based on the Australian system (National Committee on Soil and Terrain 2009). This method produced an acceptable prediction of soil drainage validated by randomly held-back data points. This method was considered somewhat incompatible with this study as knowledge of soil colour-based drainage sequences was lacking in substantial areas of the project.

Approaches similar to these were considered; however, ‘Soil Drainage Class’ was chosen to test as per the National Committee on Soil and Terrain (2009) as a training dataset to predict soil drainage class across the study area directly, without the added complexity of generating a soil–landscape model or the time and expense associated with obtaining extra covariates such as hyperspectral imagery or EMI mapping, which can be expensive to obtain over large areas. In general, the surveyors' expert knowledge (in terms of soil drainage at sampled locations) was spatially extrapolated across the project landscapes using the available covariates to explain the spatial variation between these training sites. Various modelling approaches were trialled for both categorical and continuous predictions using methods that have been successfully applied to DSM in recent literature. These included DT, regression tree (RT) and random forests (RF) approaches (see *Modelling* section).

Suitability parameters were derived by the TIA from industry and expert consultation, including the Australian Soil Drainage Class Classification standard (National Committee on Soil and Terrain 2009) used by Tasmanian growers and agronomists. Common surrogate drainage predictors include soil texture, depth to mottling, colour, topographic position, or a combination of these. Although somewhat subjective, Australian drainage classification uses these to estimate how quickly excess moisture is removed from the soil profile and landscape in conjunction with other considerations such as soil structure, porosity, water-holding capacity, water source, evapotranspiration, slope gradient and length (National Committee on Soil and Terrain 2009).

The objectives of this paper are therefore to: (*i*) test the integration of qualitative expert-based soil drainage estimates with quantitative DSM methods to produce predictive soil drainage surfaces; and (*ii*) compare DSM methods for both class and index mapping of soil drainage, and test their applicability to enterprise suitability mapping.

Materials and methods

Study areas

The Tasmanian geology generally determines soil pattern due to the strong influence of rock type upon soil formation (Spanswick and Zund 1999). The Meander study area aligns with the Meander Irrigation Scheme and it was selected to test a variety of different soils, land uses and landscapes. The area contains soils of the Launceston Tertiary Basin to the east, comprising a series of Quaternary alluvium river terraces of imperfectly to poorly drained Sodosols (Isbell 2002) (Lixisols or Solonetz; IUSS Working Group WRB 2007), and Black cracking Vertosols (Isbell 2002) (Vertisols; IUSS Working Group WRB 2007) in drainage depressions and recent flood plains. Well-drained Tertiary basalt soils (Red Ferrosols; Isbell 2002) (Nitisols or Acrisols; IUSS Working Group WRB 2007) are dominant on the hills surrounding Deloraine, and to the south, a ridge of undifferentiated conglomerate and Permian Sandstone sequences has formed shallow skeletal Rudosols (Isbell 2002) (Regosols; IUSS Working Group WRB 2007). Poorly drained, complex alluvial soils are common in the southern Dairy Plains area, previously mapped as a miscellaneous soil unit consisting of stream alluvium, marsh and swamp deposits, formed by past weathering and depositional processes from diverse surrounding lithology types (Spanswick and Zund 1999). Soil complexes consist of Hydrosols, Kandosols and Chromosols (Isbell 2002) (Gleysols, Fluvisols, and Lixisols; IUSS Working Group WRB 2007) in this area.

The Midlands project area covers 27 000 ha of the southern part of the Midlands Irrigation Scheme, an area from Oatlands and north to Tunbridge. The Tunbridge area comprises recent and higher level alluvial terraces of Sodosols (Isbell 2002) (Lixisols or Solonetz; IUSS Working Group WRB 2007), with Black cracking Vertosols (Isbell 2002) (Vertisols; IUSS Working Group WRB 2007) in drainage depressions and recent alluvial deposits. Stony Brown Dermosols (Isbell 2002) (Luvisols; IUSS Working Group WRB 2007) have formed on Jurassic dolerite hills to the south. Triassic sandstone has been capped by the dolerite on foot-slopes to the east, and alluvium has been covered by intermittent dolerite fans, adding to the spatial soil complexity. Annual rainfall is <500 mm year^{-1}, resulting in widespread sodicity (exchangeable sodium percentage >6%), with small areas of primary salinity (Kidd 2003). Triassic sandstone and Permian mudstone hills in the Oatlands vicinity have formed imperfectly to moderately well-drained Chromosols and Sodosols (Isbell 2002) (Luvisols, Lixisols, Phaeozems; IUSS Working Group WRB 2007) (Spanswick and Kidd 2001).

Enterprise suitability

Land suitability rules comprising soil, climate and landscape parameters were developed by TIA for 20 different enterprises using a four-class (well-suited, suitable, moderately suited, unsuited) most-limiting factor approach (Klingebiel and Montgomery 1961). Enterprises included a range of broad-acre and horticultural crops, with suitability parameters and ranges determined through interrogation of existing literature, TIA agricultural trials, expert advice, and formal workshops with industry representatives, agronomists and growers (Table 1). This work identified the key soil and climate parameters for each enterprise and their threshold values with respect to physical or chemical agronomic limitations. Climate surfaces were generated using digital modelling from 271 temperature sensors and terrain covariates. Input parameter and final suitability surfaces were produced at a ground resolution of 30 m. A sample drainage-suitability rule-set for blueberries is listed in Table 2 (Tasmanian Institute of Agriculture 2012).

Table 1. Enterprise suitability parameters

Soil	Climate	Terrain
pH (1 : 5 in water) (0–0.15 m)	Frost risk (seasonal, by enterprise)	Slope %
EC (1 : 5 in water) (0–0.15 m)	Mean max. monthly temp.	
Stone content (0–0.15 m)	Rainfall	
Soil depth (m)		
Clay % (0 to 0.15 m)		
Depth to sodic layer (ESP >6)		
Exchangeable calcium (0 to 0.15 m)		
Exchangeable magnesium (0 to 0.15 m)		
Soil drainage class		

Covariate data

SCORPAN environmental variables (soil covariate data) (McBratney *et al*. 2003) were compiled for both study areas to enable spatial predictions of each soil parameter. This involved co-registration of available covariate surfaces into a common mapping grid-base and generation of terrain derivatives from the DEM (Table 3).

For Meander, the existing soil map (Spanswick and Zund 1999) was partially disaggregated from original soil-association map units to predicted association components using a DT approach, and extrapolated into unmapped areas for use as a covariate. Ground-based gamma radiometric mapping was undertaken by CSIRO Land and Water to complete the partial coverage in Meander West (Viscarra Rossel *et al.* 2013). A GPS-enabled gamma radiometer recorded total count, potassium, uranium and thorium over a series of transects; the data were extrapolated into unmapped areas using terrain covariates as environmental predictors, by RF modelling (see *Modelling* section). The estimated radiometrics-terrain surfaces were then compared against an existing, overlapping radiometric-mapped area to derive a linear relationship, which was applied to the estimated surface to make it consistent with the 'true' measurement. This approach could introduce a degree of 'circularity' in using terrain for both soil and radiometric predictions and may therefore introduce some error into the DSM models; however, model training and validation metrics improved when using the estimated radiometric covariates compared with their non-use. The radiometric-terrain estimate provided a complex measure of covariance to the target variable that most modelling approaches would otherwise miss.

In general, the estimated radiometric map highlighted the complexity of the unmapped region where large, featureless, alluvial expanses had been mapped as stream alluvium, marsh and swamp deposits (Spanswick and Zund 1999), with little chance of predicting their properties based on terrain alone without proximal radiometric sensing (Viscarra Rossel *et al.* 2013). This process allowed gamma radiometrics to be used across the whole project area, and it was shown to be an important predictor of many soil properties (see *Results and discussion*).

Table 2. Enterprise suitability for soil drainage, blueberries

Suitability rating	Drainage class
Well suited	Well to moderately well
Suited	Imperfect
Moderately suited	Imperfect
Unsuited	Poor to very poor

Table 3. Spatial predictors (covariates)

SAGA GIS: System for Automated Geoscientific Analyses: http://www.saga-gis.org; DEM, digital elevation model

Scale	Spatial covariates	Scale/resolution	Reference/source
		Categorical data	
Regional	Soil map	1 : 100 000	Leamy 1961; Spanswick and Zund 1999; Spanswick and Kidd 2001
Regional	Land capability map	1 : 100 000	Noble 1993
Regional	Land-use map	1 : 50 000	DPIPWE 2012 unpubl. data
Regional	Vegetation map (TASVEG) v 2.0	1 : 25 000	DPIPWE 2009 (http://www.dpiw.tas.gov.au/tasveg)
Regional	Surface geology map	1 : 25 000	Mineral Resources Tasmania 2008 (http://www.mrt.tas.gov.au/portal/page?_pageid=35,832332&_dad=portal&_schema=PORTAL)
		Remote sensing	
Local	Rapid eye multispectral	5 m	Cradle Coast Authority 2010 (see http://blackbridge.com/rapideye/mosaics/index.html)
Local	SPOT Bands 1,2 and 3	5 m	SPOT Image 2009 (see http://www.astrium-geo.com/en/143-spot-satellite-imagery)
Local	SPOT NDVI	30 m (processed)	SAGA GIS
Local	LandSat principal components	30 m (processed)	SAGA GIS
Local	Gamma radiometrics (radioactive nuclides: K, U, Th, total dose)		Mineral Resources Tasmania 2004 http://www.mrt.tas.gov.au/portal/page?_pageid=35,832439&_dad=portal&_schema=PORTAL
		Terrain	
Local	SRTM DEM-S	30 m	1 arc-second DEM, adaptively smoothed, Geoscience Australia 2011 (http://www.ga.gov.au/meta/ANZCW0703014016.html)
Local	Slope, aspect, curvatures (plan and profile), topographic wetness index (TWI), SAGA wetness index (SWI), multi-resolution valley bottom flatness (MrVBF), multi-resolution ridge top flatness (MrRTF), northness (Sin(Aspect)), eastness (Cos(Aspect)), normalised height, slope height, vertical distance to channel network, height above channel network	30 m	SAGA GIS

Soil sampling and validation sites

A conditioned Latin Hypercube sampling design was used for an initial 20 000-ha Meander area for model training, a stratified random sampling approach based on maximally stratifying the full multivariate distribution (Minasny and McBratney 2006). An alternative, stratified random sampling approach was used for both training and validation of the remaining 50 000 ha of both areas. Fuzzy k-mean clustering of available covariates was used as stratification, where sampling sites could be spatially adjusted within cluster areas if access to an intended site was not possible (D. B. Kidd, B. P. Malone, A. B. McBratney, B. Minasny, M. Webb, unpubl. data).

Field sampling and soil analysis

Soil was sampled using a 0.05-m-diameter percussion soil corer to a depth of 1.5 m and subsampled by horizon. Cores and surrounding landscape position were described according to Australian Soil and Land Survey guidelines, including soil drainage-class estimates with corresponding drainage-class code (Table 4) (National Committee on Soil and Terrain 2009). This is a general, but expert-based, field observation that describes the soil and site drainage likely to occur in most years, and considers several both internal and external influences. Internal influences include soil structure, texture, porosity, hydraulic conductivity, moisture-holding capacity, colour and mottling, while external considerations include slope length, landscape position and likely water sources, (National Committee on Soil and Terrain 2009). The field surveyor determines a combination of the above factors to make an estimate on the soil drainage class at the site location, based on expert knowledge of the environment and the following observations as per the National Committee on Soil and Terrain (2009).

Very poorly drained soils

Very poorly drained soils are most often identified by landscape position and current moisture status. They remain wet for most of the year, and often occur in depressed areas. Soils have strong gleying throughout the profile and accumulated surface organic matter. Any or all of surface, subsurface or groundwater flow are identified as the main water sources.

Poorly drained soils

Poorly drained soils are wet for several months of the year and may be affected by perched watertables or surface ponding. Most horizons have gleyed or mottled clays close to the surface. Subsurface or groundwater is the main water source, which is supplemented by rainfall.

Table 4. Reclassification ranges of continuous soil drainage index

Drainage class code	Drainage class	Index ranges
1	Very poorly drained	<1.5
2	Poorly drained	1.5–2.5
3	Imperfectly drained	2.5–3.5
4	Moderately well-drained	3.5–4.5
5	Well-drained	4.5–5.5
6	Rapidly drained	>5.5

Imperfectly drained soils

Imperfectly drained soils occur in flatter areas and are wet for several weeks at a time. Lower horizons show mottling and rust-coloured linings of root channels. Rainfall is considered the main water source for soils with high water-storage capacity, or groundwater if water-storage capacity is low.

Moderately well-drained soils

Soils are usually medium to fine in texture (e.g. light to medium clays), with drainage impeded by a combination of lack of slope, shallow watertable, or low permeability due to structure (weakly structured soils). Soils have few or no mottles and will remain wet for up to a week after a rainfall event.

Well-drained soils

Well-drained soils are often medium in texture (e.g. clay loams, and well-structured light clays), allowing excess water to be removed by either vertical or lateral subsurface flow, and will only remain wet for several days after a rainfall event.

Rapidly drained soils

Rapidly drained soils are usually coarse-textured and/ or shallow (e.g. aeolian sands). Highly permeable layers will allow excess water to rapidly flow downwards through the profile, with rapid subsurface lateral flow on steeper slopes. These soils will only remain wet for less than a day after a rainfall event.

Data preparation

Soil observations were spatially intersected with all available covariates using SAGA GIS (System for Automated Geoscientific Analyses release 2009; http://www.saga-gis.org) to allocate individual covariate values to each drainage estimate for model training data. All covariates for each study area were spatially amalgamated into a set of values for each pixel. Model relationships (between the observations and covariate values) were applied to the combined covariates for drainage class or index value predictions at each pixel.

Modelling

Decision-tree classification of soil types and discrete properties is a popular DSM methodology that analyses and partitions covariate patterns to create predictive rules (Moran and Bui 2002; Hollingsworth *et al*. 2006; MacMillan 2008). This approach was used for soil drainage-class predictions. However, as these classes can be considered ordered by a numerical coding system, it was also decided to test whether predictions of drainage could also be made as a continuous index. With this approach, one can show gradational landscape trends and spatially display the subtle variations in soil water movement that are otherwise masked by the class thresholds. Another well-documented approach that has had success in predicting continuous soil properties is RT modelling (DTs with linear regression models at the nodes) (Moran and Bui 2002), and it was used to generate the soil drainage index. Another popular approach, RF (Cutler *et al*. 2007; Grimm *et al*. 2008; Liaw and Wiener 2002; Wiesmeier *et al*. 2011), was also tested as an alternative method for developing the drainage index; RF generates many regression trees from a random bootstrap sample, with the remaining data (called 'out-of-bag' data) used for validation of the tree. Splits are made from a random selection of covariates and based on the strongest predictors (Stum *et al*. 2010). Regression kriging, a hybridised modelling approach that incorporates regression modelling and interpolated model residuals, has been shown to improve model performance (Odeh *et al*. 1995; Hengl *et al*. 2007), and was tested to determine improvements in validation metrics. From these approaches, the best predictions in terms of statistical and field validation were selected as an input for the suitability modelling.

Decision tree modelling

The software See5© (RuleQuest Research:http://www.rulequest.com) was used to construct a series of DTs using the available covariates and sampling site descriptions to spatially predict drainage class estimates as categorical data (Table 3). Several prediction settings were trialled until the best overall training and validation relationships were obtained (see *Results and discussion*). Covariates were 'winnowed' to determine and use only the most correlated predictors for the model. The 'Rule Utility Ordering' option was also selected to use as a priority, higher in the DT process, those covariates that introduce the least amount of error into predictions, improving overall accuracy. 'Boosting' (for 10 trials) was selected in order to reduce overall error by concentrating subsequent trees on misclassified instances in the preceding classifier, and improving these in the next. Validation site-drainage estimates were used to test the DT model using confusion matrices.

Regression tree modelling

The software Cubist© (RuleQuest Research:http://www.rulequest.com) was used to construct regression tree rule-sets to spatially predict soil drainage class expressed as a continuous drainage index (Table 4), such that the class codes (1–6) were treated as a continuous variable. A five-member committee model was assembled where the first rule is constructed, and subsequent rules are formed to minimise the errors present in the previous rule-set, improving overall predictive accuracy. Due to the scarcity of poorly (Class 1) or rapidly (Class 6) drained sites in the training and validation sampling (see *Results and discussion*), the model was allowed an extrapolation of up to 20% outside the training data range to ensure that the full range of drainage conditions for both areas was covered. Regression kriging of the Cubist predictions was tested to ascertain any improvement to validation rates; however, a semi-variogram fit of RT residuals showed poor spatial correlation (no trend of variance with distance). Consequently, incorporation of the residuals as a spatial random variable by kriging did not improve validation rates and so it was not used in model outputs. Principal components of the available covariates were also tested with both the RT and DT approaches, they but did not improve model training or validation metrics.

Random forests–residual kriging (RF-RK) modelling

Drainage class estimates at each site were used to construct a continuous soil drainage index (as per the RT approach) by regression-kriging using RFs (R statistical software; Liaw and Wiener 2002; R Development Core Team 2012), with kriging of the model residuals (Odeh *et al*. 1995; Hengl *et al*. 2007). Principal components were derived for the covariates to de-correlate and reduce co-linearity (Hengl *et al*. 2004). The RF model was constructed using the principal components with the highest variable importance (determined by a stepwise linear regression) with bootstrapping using 10% of samples for 100 iterations, and 1000 regression trees (Liaw and Wiener 2002). Predictive errors showed a reasonable spatial correlation when fitted to a semi-variogram (i.e. semi-variance increased with distance to ~2 km); thus, kriging was applied to the residuals of the RF model.

Reclassification of continuous drainage indices

The continuous soil drainage index generated by the RT and RF-RK methods principally aligned to the numerical soil drainage-class system. Suitability rules were applied to this index where, for example, an enterprise that required drainage class better than imperfect had a requirement of a rating >3 applied to the index. To aid testing of the developed surfaces, values were 'reclassified' to align with discrete drainage classes (outlined in Table 4), essentially 'rounding' the index values to the nearest whole number. Classification matrices were constructed using validation sites intersected with the class surface to measure the level of agreement.

Potential surrogate—depth to mottling

To assess the feasibility of using depth to significant mottling (National Committee on Soil and Terrain 2009) as a surrogate for soil drainage class, the relationship between depth to mottling and drainage-class prediction was investigated to determine whether this single depth to mottling measurement would be sufficient to train a drainage model. The coefficient of determination was derived from a bivariate fit between depth to mottling and the corresponding drainage class estimate at each site.

Statistical validation

Drainage class observations of the validation dataset were intersected with the corresponding predictions from the RT and RF-RK models, and validation was quantified by the coefficient of determination (R^2), concordance correlation coefficient (Lin 1989), and residual standard error (RSE). Validation for the DT modelling and re-classified RT and RF-RK models was performed by intersecting the independent validation site numerical drainage class with the predicted drainage class, or re-classified predictions. Agreement statistics (kappa coefficients; Cohen 1960) and classification accuracies were generated from confusion matrices determining the proportion of successfully predicted classifications and overall model performance.

Field validation

Roadside field validation was undertaken for both the soil drainage index and soil drainage-class mapping using a GPS-enabled laptop with GIS. Mapping agreement was determined by expert assessment of visual indicators such as landscape position, vegetation, land use, management, infrastructure and post-rainfall surface ponding.

Results and discussion

Potential surrogate—depth to mottling

A bivariate fit of depth to significant mottling, as defined in the Australian Soil Survey handbook (National Committee on Soil and Terrain 2009), against soil drainage showed a poor relationship, with an R^2 of 0.12, and root mean-square error (RMSE) of 27.3. This is due to other external influences such as landscape position and soil colour and texture informing soil drainage-class estimates in the field. Consequently, methods that directly predicted soil drainage class from field site descriptions were favoured over using depth to mottling as a surrogate.

Site data

The majority of both training and validation sites for each area was classified by the surveyors as imperfectly drained (Class 3), which fits with existing soil survey drainage estimations for known soil profile classes and expert knowledge of the two areas (Leamy 1961; Spanswick and Zund 1999; Spanswick and Kidd 2001; Kidd 2003). Figures 2 and 3 show the cumulative frequency of drainage estimates for both areas.

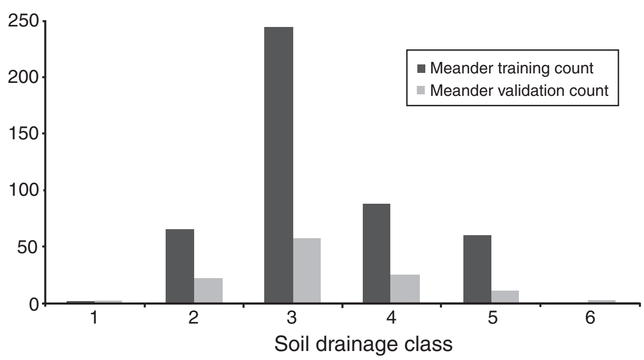

Fig. 2. Relative frequency of Meander soil drainage classes.

Modelling methods (Meander)

Decision tree

The DT modelling using available terrain, satellite and radiometric covariates showed good results. Table 5 lists the covariates used in the model trees, and the percentage of training data for which each covariate is used in predicting a class. Elevation, aspect, slope derivatives, and wetness indices were all good predictors of soil drainage, which is consistent with knowledge of Tasmanian landscapes and influence on soil drainage and moisture movement (Kidd 2003). Radioactive potassium tends to follow water-borne soil deposition zones

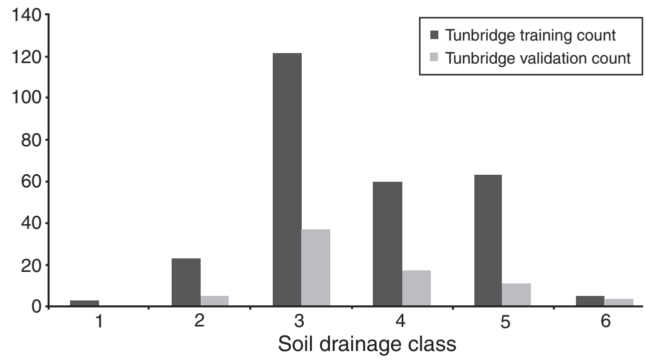

Fig. 3. Relative frequency of Tunbridge soil drainage classes.

Table 5. Decision tree (See5©) covariate usage
See Table 3 for covariate information

Covariate	Model (rule) usage (%)
SRTM DEM (m)	100
Normalised height	100
Slope height	100
Eastness	100
SWI	99
MrRTF	99
Radioactive potassium (%)	96
MrVBF	94
Northness	81
Total radiometric dose	76
Curvature class	66
Mid-slope position	63

Table 6. Decision tree soil drainage class classification matrix: Meander training and validation

Drainage class	Drainage class (no. of times classified), training, [validation] 1	2	3	4	5	6	Totals	User accuracy (%)
1	**0,[0]**	0,[0]	2,[2]	0,[0]	0,[0]	0,[0]	2,[2]	0,[0]
2	0,[0]	**65,[6]**	0,[14]	0,[1]	0,[1]	0,[0]	65,[22]	100,[27.3]
3	0,[0]	1,[6]	**243,[37]**	0,[10]	0,[5]	0,[0]	244,[58]	99.6,[63.8]
4	0,[0]	0,[1]	2,[10]	**84,[9]**	1,[4]	0,[0]	87,[24]	96.6,[37.5]
5	0,[0]	0,[0]	1,[4]	0,[1]	**59,[7]**	0,[0]	60,[12]	98.3,[58.3]
6	0,[0]	0,[0]	0,[1]	0,[0]	0,[0]	**0,[0]**	0,[1]	x,[0]
Totals	0,[0]	66,[13]	248,[68]	84,[21]	60,[17]	0,[0]	458,[119]	
Producer accuracy (%)	x,[x]	98.5, [46.2]	98.0, [54.4]	100, [42.9]	98.3, [41.2]	x,[x]		

(Viscarra Rossel *et al*. 2013), and was also correlated with drainage.

Table 6 shows the DT classification rates for each drainage class for both model training and independent validation in the Meander area. The model performed well, with the majority of classes classifying correctly. Thorough testing of the 'very poor' or 'rapidly' drained classes was not possible due to a scarcity of these conditions in the study area. The independent validation set showed that the DT model tended to classify the majority of 'poorly drained' and 'well drained' sites as 'imperfectly drained'. The 'imperfectly drained' class had the lowest misclassification rate due to the higher proportion of this class in the Meander area and the subsequent high number of training data to better construct the DT model for this class. While both producer and user accuracies were good for training, the validation user accuracy rates (the percentage of classes correctly classified) were reasonable for 'imperfectly' and 'well' drained sites but generally poor for 'poorly' and 'moderately well' drained sites (27.3% and 37.5% correctly classified samples, respectively). The overall accuracy (the total of correctly classified sites compared with the overall number of sites as a percentage) was excellent at 98.5% for training but was reduced to 49.6% for validation.

Table 7. Regression tree model covariate usage for tree partitions and tree regression models (Meander)
See Table 3 for covariate information

Covariate	Covariate usage (%) Tree partition	Model usage
Geology (25k)	40	0
Standardised height	40	57
SRTM DEM (m)	36	21
Radioactive potassium (%)	7	14
Terrain ruggedness index	2	85
Valley depth	0	72
Slope (%)	0	65
Normalised height	0	57
Slope height	0	48
SWI	0	45
Curvature	0	35
TWI	0	35
Mid-slope position	0	15
MrVBF	0	6
Profile curvature	0	6
Analytical hillshade	0	5

Regression tree

Table 7 shows the percentage usage of each covariate for both partitioning the model trees and the model usage percentage within each partition. The model was partitioned mainly using geology and elevation, and to a lesser extent radioactive potassium and terrain ruggedness, with regressions dominated by the Terrain Ruggedness Index, slope and valley depth.

RT predictions

Table 8 shows the training and validation model agreements for the Meander area. The drainage indices (treated as continuous data) include the coefficient of determination and concordance coefficient (agreement around a 1 : 1 line; Lin 1989). The classified drainage index (as discrete classes) includes the kappa coefficient and 'kappa with linear weighting' as a measure of correct classification for both model training and validation (Cohen 1960, 1968). Numerical drainage class can be considered as ordered data. Therefore, a misclassification one drainage class either side of the actual category implies a better model fit than if the data were nominal. This meets the conditions for partial credit where kappa with linear weighting can be used as a more realistic measure of classification than using an unweighted kappa coefficient (Cohen 1968). Some of the performance metrics used were not applicable for either continuous or ordinal datasets and were therefore excluded from Table 8 (denoted by 'x'). Unweighted kappa values were computed using JMP® (version 9; SAS Institute Inc., Cary, NC, USA) and linearly weighted kappa values computed using VassarStats online statistical computational software (http://www.vassarstats.net). Kappa coefficient values were applied as a generalised measure of agreement and overall model performance; however, other factors such as linear weighting for ordinal data, and class classification rates (e.g. user and producer accuracies), should be considered (Fleiss *et al*. 2004).

Table 9 shows the agreement rates for the re-classified RT soil drainage predictions. As with the DT approach, the re-classified RT training classification rates were generally good, with a user accuracy >80% for all classes other than 'moderately

Table 8. Summary of Meander soil drainage-class spatial modelling metrics
x, not applicable

Method/metrics	R^2	Concordance coefficient	Residual standard error	Kappa	Kappa (linear weighting)	Overall classification accuracy (%)
DT training	x	x	x	0.79	0.86	98.5
DT validation	x	x	x	0.22	0.32	49.6
RT training	0.79	0.86	0.27	x	x	x
RT validation	0.39	0.57	0.46	x	x	x
Classified RT training	x	x	x	0.63	0.72	77.3
Classified RT validation	x	x	x	0.27	0.37	54.6
RF-RK training	0.90	0.91	0.21	x	x	x
RF-RK validation	0.36	0.48	0.39	x	x	x
Classified RF-RK training	x	x	x	0.79	0.84	86.9
Classified RF-RK validation	x	x	x	0.15	0.29	48.7

well-drained'; 42% of 'poorly drained' sites were classified as 'imperfectly drained' and 47% of 'well drained' sites were classified as 'moderately well-drained'. Although not performing as well for the training classification rates as the DT approach, the re-classified RT performed better overall when tested against the independent validation sites, with a linearly weighted kappa of 0.37, compared with 0.32 for the DT model (Table 8), a fair to moderate validation agreement (Altman 1991). For validation of drainage classes 2, 3, 4 and 5, the DT approach had a user accuracy rate of 27, 64, 38 and 58%, respectively, whereas the RT approach classified at 23, 74, 58 and 25% for user accuracy. The overall accuracy for training was slightly lower than the DT approach (77.3%), while the validation overall accuracy was 54.6%, an improvement over the DT validation.

Random forests

Table 10 highlights the agreement metrics for the classified RF-RK for training and validation respectively. This approach showed the highest kappa with linear weighting agreement (0.84) for model training. As with the RT agreement matrix, training classification rates were generally good for poorly and well-drained classes (2 and 5, respectively) but tended to under-classify moderately well-drained sites, with a 69.3% user accuracy rate. However, as demonstrated with the RF-RK drainage index that tended to over-fit the training data, the lowest validation rate was obtained for the classified surface with a weighted kappa of 0.29 (Table 8). Table 10 shows generally poor user accuracy validation rates for all classes other than class 5 (well-drained), and failed to correctly classify any poorly drained (class 2) sites. Overall accuracy was 86.9% for training and 48.7% for validation.

Table 8 summarises the training and validation rates for all approaches. In comparing observed *v.* predicted values for soil drainage, RF-RK showed the best training agreement in terms of coefficient of determination and concordance (Lin 1989), with 0.90 and 0.91, respectively. RT also showed good concordance, with 0.79 and 0.86, respectively. Residual standard error was close for both methods. However, the RT methodology showed a better validation than the RF-RK model, with an R^2 of 0.39–0.36, and concordance of 0.57–0.48, respectively. This implies that the RF-RK approach tended to 'over-fit' the training data, with the more substantial discrepancy between the training and validation rates than the RT approach. The discrepancy between concordance and R^2 also implies that both approaches tended to under-predict drainage for better drained sites. Residual standard error was again close for validation of both approaches, with 0.46 for RT and 0.39 for RF-RK. This is less than half of a soil drainage class in terms of an index, which is considered acceptable for the regional-resolution suitability mapping requirements.

Tunbridge modelling

DT predictions

In the Tunbridge region, the DT approach used the existing soil mapping for all trees and also relied on the broad-scale geology and the topographic wetness index, but did not select any of the radiometric surfaces. The model produced a 'good' training agreement of 0.77 for the weighted kappa coefficient,

Table 9. Regression tree soil drainage class classification matrix: Meander training and validation

Site drainage class	Drainage class (no. of times classified), training, [validation] 1	2	3	4	5	6	Totals	User accuracy (%)
1	**0,[0]**	2,[1]	0,[1]	0,[0]	0,[0]	0,[0]	2,[2]	x,[0]
2	0,[0]	**38,[5]**	27,[17]	0,[0]	0,[0]	0,[0]	65,[22]	88.4,[22.3]
3	0,[0]	3,[2]	**221,[43]**	19,[12]	1,[1]	0,[0]	244,[58]	82.2,[74.2]
4	0,[0]	0,[1]	21,[9]	**63,[14]**	3,[0]	0,[0]	87,[24]	57.3,[58.3]
5	0,[0]	0,[0]	0,[3]	28,[6]	**32,[3]**	0,[0]	60,[12]	88.9,[25.0]
6	0,[0]	0,[0]	0,[1]	0,[0]	0,[0]	**0,[0]**	0,[1]	x,[0]
Totals	0,[0]	43,[9]	269,[74]	110,[32]	36,[4]	0,[0]	458,[119]	
Producer accuracy (%)	0,[0]	58.5, [55.6]	90.6, [58.1]	72.4, [43.8]	53.3, [75.0]	x,[x]		

Table 10. Random forests soil drainage class classification matrix: Meander training and validation

Site drainage class	Drainage class (no. of times classified), training, [validation] 1	2	3	4	5	6	Totals	User accuracy %
1	**0,[0]**	2,[1]	0,[1]	0,[0]	0,[0]	0,[0]	2,[2]	x,[x]
2	0,[0]	**51,[0]**	14,[22]	0,[0]	0,[0]	0,[0]	65,[22]	96.3,[0]
3	0,[0]	0,[0]	**237,[45]**	7,[12]	0,[1]	0,[0]	244,[58]	91.2,[54.2]
4	0,[0]	0,[0]	9,[13]	**78,[11]**	0,[0]	0,[0]	87,[24]	69.3,[34.4]
5	0,[0]	0,[0]	0,[2]	28,[8]	**32,[2]**	0,[0]	60,[12]	100.0,[66.7]
6	0,[0]	0,[0]	0,[0]	0,[1]	0,[0]	0,[0]	0,[1]	x,[x]
Totals	0,[0]	53,[1]	260,[83]	113,[32]	32,[3]	0,[0]	458,[119]	
Producer accuracy (%)	0,[0]	78.5, [0]	97.1, [77.6]	69.0, [45.8]	100.0, [16.7]	x,[0]		

Table 11. Summary of Tunbridge soil drainage-class spatial modelling metrics
x, not applicable

Method/metrics	R^2	Concordance coefficient	Residual standard error	Kappa	Kappa (linear weighting)	Overall classification accuracy (%)
DT training	x	x	x	0.76	0.77	83.8
DT validation	x	x	x	0.09	0.13	39.5
RT training	0.43	0.57	0.48	x	x	x
RT validation	0.34	0.53	0.50	x	x	x
Classified RT training	x	x	x	0.33	0.48	53.3
Classified RT validation	x	x	x	0.31	0.47	52.6

but produced a poor weighted kappa validation agreement of 0.13 (Table 11), again indicating a model over-fit of training data. Overall accuracy for training was 83.8% but only 39.5% for validation.

RT soil drainage predictions

The RT approach used more available covariates, only applying the existing soil mapping to partition the model trees. Radiometric surfaces were used in conjunction with the terrain and, to a lesser extent, satellite vegetation indices. The RT model produced reasonable training metrics, with R^2 and concordance of 0.43 and 0.57, respectively. Independent validation was also reasonable, with a concordance of 0.53 and standard error of 0.5 (Table 11). The re-classification of the RT soil drainage index into discrete soil classes showed a 'moderate' validation, with a weighted kappa value of 0.47 (Table 11). Despite overall classification accuracy for training of 53.3% (which was significantly lower than the DT agreement rate), a validation accuracy of 52.6% was a substantial improvement over the DT validation, implying a more realistic model (Table 11).

Model comparisons

The strong relationship between soil drainage and landscape in Tasmania explains the good spatial correlation with the available covariates for all modelling approaches. Although the DT discrete model worked reasonably well in Meander, poor validation was attained in the Tunbridge area. Similarly, the RF-RK model for Meander tended to over-fit the training data at the expense of independent validation; that is, the model developed spatial relationships that closely associated the available predictors between the training points, but did not necessarily reflect what is actually occurring in the landscape, resulting in a higher rate of unexplained variability. Hence, the RT approaches for soil drainage predictions in both areas were chosen as suitability inputs, as good validation outputs were achieved compared with the DT and RF-RK models for both the generated soil drainage index, and re-classified, discrete soil drainage-class mapping.

The expert knowledge of the drainage at a particular location was effectively captured and extrapolated to surrounding landscapes based on the spatial correlation of available predictors. However, this training might be considered somewhat subjective and it relies on the surveyor's experience and local soil–landscape knowledge. There may also be some uncertainty over the linearity of the Australian Soil Drainage Class system. Despite this, the covariates were spatially correlated with the site estimates, and acceptable model validation was achieved through the RT approach, despite training data being generated by two different surveyors. Any subjectivity or discrepancies between surveyors may have been moderated by the fact that estimated classes encompass a range of slightly different drainage rates within each category.

A definitive conclusion as to the best modelling approach would not be possible without some form of quantitative evaluation of the modelled surfaces. Quantitative soil drainage analysis in the form of replicated hydraulic conductivity measurements would provide a more rigorous validation of the different modelling approaches. However, most quantitative methods tend to provide a measure of soil permeability only, without consideration or measurement of the other environmental factors that contribute to the Australian drainage class estimation. Such quantitative measurements would therefore only provide a surrogate measure for validation. The time and cost associated with these replicated measurements was not within the resources of the present study.

An advantage of the RT approach is that both continuous drainage indices and discrete drainage class mapping can be produced. The drainage index is 'visually appealing' and demonstrates how soil drainage spatially trends with landscape position, rather than more polygonal (and spatially unrealistic) drainage class cut-offs. Another advantage is that it can be applied to any legacy soil data that have a drainage-class estimate attached, to either derive new mapping, or improve existing mapping as a continuous, statistically validated index. It also reduces the need to apply complex hydrographical modelling functions in areas where no groundwater data or soil–landscape sequence drainage knowledge exists. The RT outputs (specifically Cubist) are easier to interpret than the RF approach, where rule-sets can be generated to show how each covariate is used within predictions, partitioning of these data into discrete spatial covariate zones, and the regression relationships within.

The soil drainage index that was generated aligned well with expert drainage knowledge, conceptual soil–landscape patterns of area, and existing Soil Profile Class definitions (Spanswick and Zund 1999; Spanswick and Kidd 2001; Kidd 2003). Importantly, in addition to the acceptable statistical validation of results, field validation was positive, where mapping aligned with the visual road-side indicators described in the methodology. Drainage map and suitability samples are shown in Fig. 4.

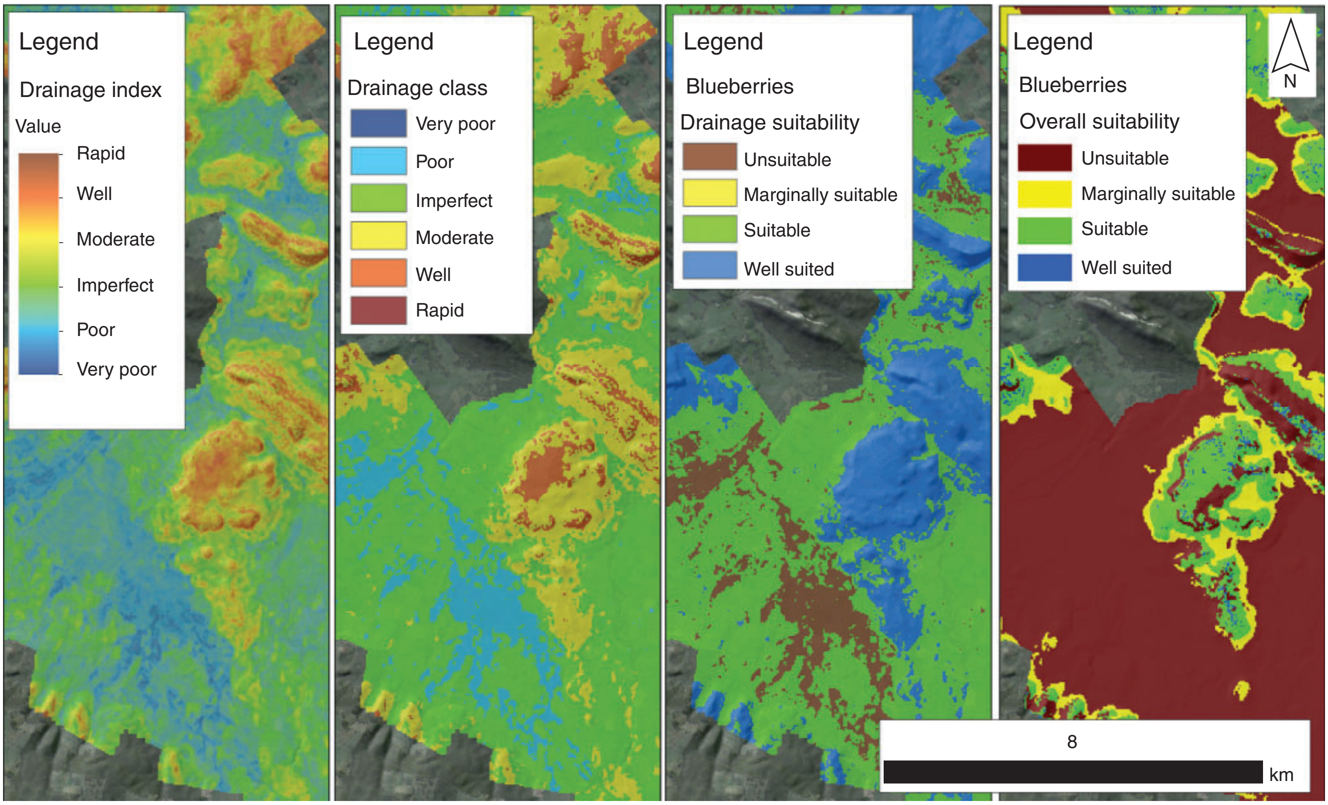

Fig. 4. Partial extent—Meander drainage index, classified drainage class, blueberries drainage suitability and blueberries overall suitability.

The index was used for the suitability predictions with drainage class cut-offs applied for each suitability class. Twenty enterprise suitability surfaces (at 30-m resolution) were generated using all digital soil surfaces, with the limiting factors for each pixel attached as potential management constraints. In the example of blueberries, soil drainage was a limitation to suitability in substantial areas, demonstrating the importance of this mapping (Fig. 4). All surfaces are available on the Tasmanian Government spatial web-based portal for public access (www.theLIST.tas.gov.au).

Conclusions

The drainage surfaces generated using the regression-tree (RT) approach aligned well with expected landscape drainage patterns, known soil profile classes, and visual field indicators, and produced acceptable statistical agreement with independent qualitative validation data. Compared with the decision-tree and random-forests modelling, RT produced better results when validated both as a drainage index, and as a discrete reclassified surface within this study. In summary, the RT approach had some limitations but was generally acceptable for the purposes of this project, the positives including:

- Expert soil drainage knowledge is extrapolated across the landscape.
- The surfaces are consistent with Australian standards and industry terminology.
- It is relatively rapid to implement (compared with replicated soil physical drainage measurements).
- Drainage Class codes can be used to derive a drainage index, which is visually appealing and a good indication of how drainage gradually changes with respect to landscape.
- The approach can be applied to legacy data, potentially improving existing mapping, and generating new soil drainage indices.

Negatives of this approach include:

- The mapping requires recruitment of experienced soil surveyors to implement field work.
- Drainage estimates are somewhat subjective, which could introduce some inconsistency between datasets from multiple surveyors, and potential modelling errors.

The generated drainage index surfaces were found to be an important input parameter for the operational government-enterprise suitability mapping. The approach proved a viable technique for spatial drainage class predictions within available project resources, for regional-resolution operational digital soil assessment.

Acknowledgements

The authors acknowledge the following: Australian Research Council Linkage Project Funding; Chris Grose and Rob Moreton for soil profile and drainage descriptions; Regan Parkinson and Jo Wang for field assistance and sample processing; Leigh Sparrow, Bill Cotching, Fiona Kerslake and Rowan Smith (Tasmanian Institute of Agriculture (TIA)) for development of the suitability rule-sets; Rhys Stickler, Peter Voller and Ashley Bastock for project management; and the farmers of the project areas for their cooperation and access.

References

Altman DG (1991) 'Practical statistics for medical research.' (Chapman & Hall: London)

Bell JC, Cunningham RL, Havens MW (1992) Calibration and validation of a soil–landscape model for predicting soil drainage class. *Soil Science Society of America Journal* **56**, 1860–1866. doi:10.2136/sssaj1992.03615995005600060035x

Bell JC, Cunningham RL, Havens MW (1994) Soil drainage class probability mapping using a soil–landscape model. *Soil Science Society of America Journal* **58**, 464–470. doi:10.2136/sssaj1994.03615995005800020031x

Cohen JA (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* **20**, 37–46. doi:10.1177/001316446002000104

Cohen JA (1968) Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin* **70**, 213–220. doi:10.1037/h0026256

Cotching WE, Kidd DB (2010) Soil quality evaluation and the interaction with land use and soil order in Tasmania, Australia. *Agriculture, Ecosystems & Environment* **137**, 358–366. doi:10.1016/j.agee.2010.03.006

Cutler DR, Edwards TC, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ (2007) Random forests for classification in ecology. *Ecology* **88**, 2783–2792. doi:10.1890/07-0539.1

Fleiss JL, Levin B, Paik MC (2004) The measurement of interrater agreement. In 'Statistical methods for rates and proportions'. pp. 598–626. (John Wiley & Sons, Inc.: New York)

Grimm R, Behrens T, Märker M, Elsenbeer H (2008) Soil organic carbon concentrations and stocks on Barro Colorado Island—Digital soil mapping using Random Forests analysis. *Geoderma* **146**, 102–113. doi:10.1016/j.geoderma.2008.05.008

Grunwald S (2010) Current state of digital soil mapping and what is next. In 'Digital soil mapping'. (Eds J Boettinger, D Howell, A Moore, A Hartemink, S Kienast-Brown) pp. 3–12. (Springer: Dordrecht, the Netherlands)

Hengl T, Heuvelink GBM, Stein A (2004) A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma* **120**, 75–93. doi:10.1016/j.geoderma.2003.08.018

Hengl T, Heuvelink GBM, Rossiter DG (2007) About regression-kriging: From equations to case studies. *Computers & Geosciences* **33**, 1301–1315. doi:10.1016/j.cageo.2007.05.001

Hollingsworth ID, Bui EN, Odeh IOA, McLeod P (2006) Rule-based land unit mapping of the Tiwi Islands, Northern Territory, Australia. In 'Developments in soil science'. (Eds AB McBratney, P. Lagacherie, M Voltz) Ch. 29, pp. 401–621. (Elsevier: Amsterdam)

Isbell R (2002) 'The Australian Soil Classification.' Revised edn. Australian Soil and Land Survey Handbooks Series 4. (CSIRO Publishing: Melbourne)

IUSS Working Group WRB (2007) 'World Reference Base for Soil Resources 2006.' First update 2007. World Soil Resources Report No. 103. (FAO: Rome)

Kidd DB (2003) Land degradation and salinity risk investigation in the Tunbridge District, Tasmanian Midlands. Department of Primary Industries, Parks, Water and the Environment. Hobart, Tas.

Kidd DB, Webb MA, *et al*. (2012) Digital soil assessment: Guiding irrigation expansion in Tasmania, Australia. In 'Digital soil assessments and beyond: Proceedings of the 5th Global Workshop on Digital Soil Mapping 2012'. Sydney. (Eds B Minasny, AB McBratney, I. Wheeler, B.P. Malone, D. van der Linden) (Taylor & Francis: London)

Klingebiel AA, Montgomery PH (1961) 'Land capability classification.' Agriculture Handbook No. 210. (USDA Soil Conservation Service: Washington, DC)

Leamy ML (1961) 'Reconnaissance soil map of Tasmania. Sheet 61—Interlaken.' (CSIRO: Melbourne)

Lemercier B, Lacoste M, Loum M, Walter C (2012) Extrapolation at regional scale of local soil knowledge using boosted classification trees: A two-step approach. *Geoderma* **171–172**, 75–84.

Liaw A, Wiener M (2002) Classification and regression by randomForest. *R News* **2**, 18–22. Available at: http://cran.r-project.org/doc/Rnews/Rnews_2002-3.pdf [verified 20 February 2014].

Lin LIK (1989) A concordance correlation coefficient to evaluate reproducibility. *Biometrics* **45**, 255–268. doi:10.2307/2532051

Liu J, Pattey E, Nolin MC, Miller JR, Ka O (2008) Mapping within-field soil drainage using remote sensing, DEM and apparent soil electrical conductivity. *Geoderma* **143**, 261–272. doi:10.1016/j.geoderma.2007.11.011

MacMillan RA (2008) Experiences with applied DSM: Protocol, availability, quality and capacity building. In 'Digital soil mapping with limited data'. (Eds A Hartemink, A McBratney, ML Mendonça-Santos) pp. 113–135. (Springer: Dordrecht, the Netherlands)

Malone BP, McBratney AB, Minasny B (2012) Digital mapping of a soil drainage index in the Lower Hunter Valley, NSW. In 'Proceedings of the Joint Soil Science Australia and New Zealand Society of Soil Science Conference'. 2–7 December 2012, Hobart, Tas. (Ed. L Sparrow) (SSA/NZSSS)

McBratney AB, Mendonça Santos ML, Minasny B (2003) On digital soil mapping. *Geoderma* **117**, 3–52. doi:10.1016/S0016-7061(03)00223-4

McKenzie N, Coughlan K, Cresswell H (Eds) (2002) 'Soil physical measurement and interpretation for land evaluation.' (CSIRO Publishing: Melbourne)

Minasny B, McBratney AB (2006) Latin Hypercube sampling as a tool for digital soil mapping. In 'Developments in soil science'. (Eds AB McBratney, P Lagacherie, M Voltz) Ch. 12, pp. 153–606. (Elsevier: Amsterdam)

Moran CJ, Bui EN (2002) Spatial data mining for enhanced soil map modelling. *International Journal of Geographical Information Science* **16**, 533–549. doi:10.1080/13658810210138715

National Committee on Soil and Terrain (2009) 'Australian soil and land survey field handbook.' 3rd edn. Australian Soil and Land Survey Handbooks Series 1. (CSIRO Publishing: Melbourne)

Niang MA, Nolin M, Bernier M, Perron I (2012) Digital mapping of soil drainage classes using multitemporal RADARSAT-1 and ASTER images and soil survey data. *Applied and Environmental Soil Science* **2012**, Article ID 430347.

Noble K (1993) Land capability survey of Tasmania. Meander Report and accompanying 1 : 100,000 Scale Map. Department of Primary Industries, Parks, Water and the Environment, Hobart, Tas.

Odeh IOA, McBratney AB, Chittleborough DJ (1995) Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. *Geoderma* **67**, 215–226. doi:10.1016/0016-7061(95)00007-B

Peng W, Wheeler DB, Bell JC, Krusemark MG (2003) Delineating patterns of soil drainage class on bare soils using remote sensing analyses. *Geoderma* **115**, 261–279. doi:10.1016/S0016-7061(03)00066-1

R Development Core Team (2012) 'R: A language and environment for statistical computing.' (R Foundation for Statistical Computing: Vienna)

Spanswick S, Kidd D (2001) Oatlands Soil Report—Revised edition. Reconnaissance Soil Map Series of Tasmania. Department of Primary Industries, Parks, Water and the Environment. Hobart, Tas.

Spanswick S, Zund P (1999) Quamby Soil Report—Revised edition. Reconnaissance Soil Map Series of Tasmania. Department of Primary Industries, Parks, Water and the Environment, Tasmania, Hobart, Tas..

Stum AK, Boettinger JL, White MA, Ramsey RD (2010) Random forests applied as a soil spatial predictive model in arid Utah. In 'Digital soil mapping'. (Eds J Boettinger, D Howell, A Moore, A Hartemink, S Kienast-Brown) pp. 179–190. (Springer: Dordrecht, the Netherlands)

Tasmanian Institute of Agriculture (2012) Enterprise Suitability Maps. http://www.dpipwe.tas.gov.au/inter.nsf/WebPages/JBAS-95335X?open

Viscarra Rossel R, Webster R, Kidd DB (2013) Mapping gamma radiation and its uncertainty from weathering products in a Tasmanian landscape with a proximal sensor and random-forest kriging. *Earth Surface Process and Landforms* doi:10.1002/esp.3476

Wiesmeier M, Barthold F, Blank B, Kögel-Knabner I (2011) Digital mapping of soil organic matter stocks using Random Forest modeling in a semi-arid steppe ecosystem. *Plant and Soil* **340**, 7–24. doi:10.1007/s11104-010-0425-z